This is an author generated pre-edited copy. The definitive version of this article is: Emma Cave, Søren Holm. Milgram and Tuskegee—Paradigm Research Projects in Bioethics. Health Care Analysis March 2003, Volume 11, Issue 1, pp 27-40. It is accessible at http://link.springer.com/article/10.1023/A%3A1025333912720

Authors:

- Emma Cave, Department of Law, University of Leeds
- Søren Holm Institute of Medicine, Law and Bioethics, University of Manchester.

Title: Milgram and Tuskegee—Paradigm Research Projects in Bioethics

(2003) 11(1) Health Care Analysis 27-40.

Abstract

This paper discusses the use of the Milgram obedience experiments and the Tuskegee syphilis study in the bioethical literature. The two studies are presented and a variety of uses of them identified and discussed. It is argued that the use of these studies as paradigms of problematic research relies on a reduction of their complexity. What is discussed is thus often constructions of these studies that are closer to hypothetical examples than to the real studies.

Conscience and morality, Milgram teaches us, do not run in smooth, logical grooves but recede and advance according to the expectations and possibilities embedded in situations both more public and politicized than hypothetical moral dilemmas (Damico 1982, p. 427)

INTRODUCTION

One of the ways in which research projects and their findings are used in ethical discussion is as paradigms. Certain studies are assumed to be so well known that you just have to mention their name, and a whole host of associations are supposedly established in the head of your reader.

Two of the most central paradigms in this sense in the field of research ethics are the Milgram obedience experiments and the Tuskegee syphilis study (Bowman, 1999, Fairchild, 1999).[1] In this paper we will present an analysis of how these studies are used in writing and teaching about research ethics. We will show that the studies are often pared down to the bare essentials when they are presented, and that the considerable complexity of the actual studies has a tendency to disappear. The first part of the paper will briefly introduce the two studies. The second and main part will then look at a variety of uses of the studies.

THE MILGRAM OBEDIENCE EXPERIMENTS

From the moment Stanley Milgram published his first findings in 1963 showing that 65% of a group of ordinary experimental subjects where willing to give another person electrical shocks up to 450 volts if prompted to do so by an experimenter the Milgram experiments have been widely discussed (Milgram, 1963). Some have subjected the experiments, and Milgram personally to severe criticism, whereas others have found the results disturbing but at the same time extremely enlightening.[2]

---

[1] This paradigmatic status can for Instance be demonstrated by doing an Internet search with the keywords “Milgram and deception” or “Tuskegee and research.”

[2] positively biased, but nevertheless very thorough and readable overview of the first 20 years of the

Although Milgram only reported one experiment in his 1963 article the whole series of experiments actually involved 18 distinct experiments with a total of approximately 1000 subjects carried out at Yale University in 1960–1963. The first full account of all 18 experiments appeared in Milgram's 1974 book "Obedience to Authority—An Experimental View" (Milgram, 1975).

The basic research question for Milgram was not stated clearly in the early works, but a plausible reconstruction is "If X tells Y to hurt Z, under what conditions will Y carry out the command of X and under what conditions will he refuse?" In order to study this question experimentally Milgram created the following experimental design:

> "Two people come to a psychology laboratory to take part in a study of memory and learning. One of them is designated as a "teacher" and the other a "learner." The experimenter explains that the study is concerned with the effects of punishment on learning. The learner is conducted into a room, seated in a chair, his arms strapped to prevent excessive movement, and an electrode attached to his wrist. He is told that he is to learn a list of word pairs; whenever he makes an error, he will receive electric shocks of increasing intensity. The real focus of the experiment is the teacher. After watching the learner being strapped into place, he is taken into the main experimental room and seated before an impressive shock generator. Its main feature is a horizontal line of thirty switches, ranging from 15 volts to 450 volts, in 15-volt increments. There are also verbal designations which range from SLIGHT SHOCK to DANGER—SEVERE SHOCK. The teacher is told that he is to administer the learning test to the man in the other room. When the learner responds correctly, the teacher moves on to the next item; when the other man gives an incorrect answer, the teacher is to give him an electric shock. He is to start at the lowest shock level (15 volts) and to increase the level each time the man makes an error, going through 30 volts, 45 volts, and so on.
>
> The "teacher" is a genuinely naive subject who has come to the laboratory to participate in an experiment. The learner, or victim *(sic!)*, is an actor who actually receives no shock at all. The point of the experiment is to see how far a person will proceed in a concrete and measurable situation in which he is ordered to inflict increasing pain on a protesting victim. At what point will the subject refuse to obey the experimenter?" (Milgram, 1974, p. 3–4).

The 18 experiments are all variations on this basic paradigm, varying the proximity of the learner, the learners pain behaviour, the proximity of the experimenter, the displayed "experience or competence" of the experimenter, the position of the teacher in the chain of persons involved in administering the shock, and the presence of rebelling peers.

The proportion of subjects defying the experimenter sometime before giving the 450 volts shock varies between 7% (in a variation where the teacher only says whether the answer given is right or wrong and someone else administers the shock) and 100% (in a variation where two experimenters disagree about whether the experiment should continue). This enormous difference in obedience in different contexts is often completely forgotten when the experiments are used in ethical discussion.

In many of the experiments the subjects displayed severe psychological distress, and this is described extensively in Milgram's articles:

> In the first four conditions 71 of the 160 subjects showed definite signs of nervous laughter and smiling. The laughter seemed entirely out of place, even bizarre. Full-blown uncontrollable seizures were observed for 15 of these subjects. On one occasion we observed a seizure so violently convulsive that it was necessary to call a halt to the experiment. (Milgram, 1965, p. 68).
>
> I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within 20 minutes he was reduced to a twitching, stuttering wreck, who was

---

discussion can be found in Miller, 1986.

> rapidly approaching a point of nervous collapse (Milgram, 1963, p. 377).

After the experiment each subject was extensively debriefed. It was explained that no electrical shocks had been given to the learner, and subjects were given an explanation of the study aimed at bolstering their self-esteem and supporting their course of action. They were later sent a report which again explained the procedures used and the findings. Finally they were sent a questionnaire and 84 percent indicated that they were glad or very glad to have participated in the study (Milgram, 1964).

Milgram's results have been extensively replicated and extended, and they seem to hold across cultures and with different types of experimental subjects (Mantell, 1971). Most of the replications are from the 1960s and 1970s and it is thus uncertain to what degree the exact levels of obedience found still holds today.

THE TUSKEGEE STUDY

The Tuskegee study is considerably older in its inception than the Milgram experiments, but it entered the debate later. From 1932–1972 the US Public Health Service conducted an experiment involving the observation of the course of untreated (and under-treated) late syphilis in 400 African American share croppers in Alabama. The intention was to compare the prognosis of the untreated syphilis participants with that of another, non-syphilitic population.

Syphilis was historically thought to be the product of sin and treatment was left to the individual sufferer. However, in the 1920s the public health system began to operate and efforts were made to treat rather than punish sufferers, partly as an attempt to stop the spreading of this infectious disease. In 1932 the best available treatment for syphilis was thought to be the heavy metal containing the drug "Salvarsan."[3] However, it was noted that, particularly in relation to late syphilis, the symptoms of the disease were perhaps less severe than the side effects of treatment. A Norwegian study that began in 1891 gave ground to this hypothesis. Thus the rationale behind the Tuskegee study at the time of the inception of the study was probably valid. The population in question had an extremely high rate of syphilis. They were underprivileged, and had therefore received little or no treatment. This was an important factor if the study was to reveal scientifically useful data. The main ethical dilemmas arise after the inception of the study and centre on three issues informed consent, exploitation, and access to treatment (see more below). It is these ethical dilemmas that have made the Tuskegee study (in)famous.

MILGRAM, TUSKEGEE AND RESEARCH ETHICS

How have the Milgram and Tuskegee studies then been used in discussions about research ethics?

Milgram's experiments have served two different purposes in writings about research ethics:

1. As a touchstone for discussions of the ethics of deceptive psychological research.
2. As a bad example of harm to research subjects.

Deceptive Research

The first paper on the ethics of Milgram's research appeared only one year after the publication of Milgram's 1963 article. In this article two major problems in Milgram's design was identified and very harshly criticised. These were 1) the

[3] It is interesting to note that the name implies that this drug could either save the sufferer from the insanity which is a common consequence of late stage syphilis or save their health.

use of deception, and 2) the willingness to let the research subjects experience quite severe emotional distress (Baumrind, 1964).

In the 1960s the whole field of research ethics was still in its infancy, and the present extensive public regulation was not in place. Deception in psychological experiments was widespread and Milgram's experiments do not stand out as exceptionally deceptive compared with other experiments of the same period (Korn, 1997). What led to the criticism of Milgram was probably a combination of the high visibility of the studies, their unexpected results, his negative moral assessment of his research subjects, and the degree of psychological distress experienced by the subjects.

Be that as it may, it is interesting that of all the deceptive experiments performed then and since a very simplified version of the Milgram experiments have attained the status as the classical example of unethical deception. Little note is taken of Milgram's debriefing efforts, if they are even mentioned.
Using the Milgram experiments as the paradigm of deceptive research may have unfortunate side-effects, because it makes it less likely that the discussion will touch upon 1) the difference between research where deception is necessary and research where deception is merely convenient, and 2) research where deception is combined with the use of institutional coercion (for many examples of this latter sort of research see Advisory Committee on Human Radiation Experiments, 1996).

Harm to Research Subjects
As described above many of the research subjects in the Milgram studies experienced a significant level of distress during the experiments, and the second line of attack in the debate about the ethics of the experiments focus on this distress. Was Milgram justified in exposing people to such distress? And can researchers in general be justified in harming research subjects?

This is the other main debate where the Milgram experiment is used as a paradigmatic example, but in this context not an example of deception but of problematic harm to research subjects.

The harm done to the subjects was foreseeable. Prior to the actual experimental series Milgram had performed pilot experiments and therefore knew that high levels of distress would occur.

Milgram and others have defended this infliction of harm in a number of ways:
1. The harm was only temporary.
2. The subjects consented to or accepted the harm subsequently.
3. The harm was outweighed by benefits to the subjects.
4. The harm was outweighed by the importance of the knowledge generated.

There seems to be no way of denying that many of the subjects in Milgram's experiments were harmed, in the same way that somebody on whom I inflict a brief pain is harmed, but it has been claimed that this harm is unimportant because it is not permanent, or at least not long lasting. There is still discussion about whether some of Milgram's subjects were actually permanently harmed, for instance by realising some very difficult truths about themselves. But, even if it can be shown that the harm was only temporary in all instances, it seems to be a very weak basis on which to justify the infliction of this harm. A harm is a harm, and even small, temporary harms need some justification.

In some of his writing Milgram claims that the acceptance of the subjects is the justification for the experiments:

> To my mind, the central moral justification for allowing my experiment is that it was judged acceptable by those who took part in it. Criticism that does not take account of the tolerant reaction of the participants has always seemed to me hollow. (Milgram, 1977, p. 21).

As mentioned above, most of his subjects were glad that they had participated in the research, and other answers from the questionnaires showed that most also found the research important. It is a commonplace in the modern discussion of research ethics that retrospective consent is not consent, but at the most acquiescence. Milgram could, however, be read not as saying that the subjects retrospectively consented, but that their evaluation of the merits of the experiment should be given substantial weight when the ethics of the experiment is assessed. The merits of this argument are questionable. Even if we discount the obvious biases that are possibly influencing the subjects' retrospective assessment of the experiment (wanting to make sense of a bad experience etc.), and take the answers to Milgram's questionnaire at face value, we still have to remember that some subjects were not glad that they had participated in the experiment. For some the harm they had experienced was still unacceptable.

A third way of justifying the foreseeable experimental infliction of psychological distress would be in terms of benefits to the subjects themselves. If for instance the psychological insights they gain about themselves are valuable insights then they may counterbalance the distress. There is unfortunately no research which can illuminate the empirical merits of this argument. A further problem is that unless it can be shown that every subject has benefited, or at least not experienced net harm, the argument collapses into a general consequentialist argument justifying the harm to some by the benefits to others.

The final way of justifying the harm caused in Milgram's experiments is by their general contribution to knowledge. The negative consequences to the subjects are outweighed by the positive benefits to all of us by having this knowledge. The results of Milgram's experiments are surprising, and they arguably tell us something about human nature which it is very important to know (Kaufman, 1967). This is probably the strongest of the possible defences of Milgram's studies, but it is a defence that comes with a price. Milgram's subjects had not consented to being harmed, and if the harm caused in this context can be justified by the good consequences generated the same must apply in many other circumstances, including many biomedical experiments. Milgram is aware of the problem and tries to block this consequence of the argument:

> I believe that it is extremely important to make a distinction between biomedical interventions and those that are of a purely psychological character, particularly the type of experiment I have been discussing. Intervention at the biological level prima facie places a subject "at risk." The ingestion of a minute dose of a chemical or the infliction of a tiny surgical incision has the potential to traumatize a subject. In contrast, in all of the social psychology experiments that have been carried out, there is no demonstrated case of resulting trauma. And there is no evidence whatsoever that when an individual makes a choice in a laboratory situation—even the difficult choices posed by the conformity or obedience experiments—any trauma, injury, or diminution of well-being results. (Milgram, 1977, p. 22).

Given Milgram's own descriptions of the psychological distress of his subjects it seems extraordinary that he feels able to claim that no trauma or diminution of well-being has ever resulted from his experiments. For most people " . . . a seizure so violently convulsive that it was necessary to call a halt to the experiment." (Milgram, 1965, p. 68) would qualify as trauma and/or diminution of well-being. The further implications of a consequentialist justification of the harm caused to unconsenting experimental subjects in the Milgram studies can therefore not be blocked in this way.

It is evident from the discussion that the Milgram experiments do have something to teach us about the acceptability of harms to experimental subjects, but using them as the paradigm case may be problematic. In the Milgram experiments the subjects were deceived and did not consent to being harmed. In the large majority of research taking place today the subjects have given their informed consent, and if the information has been sufficient they have therefore also consented to the potential harms described in the information. In this situation the central question about harm is completely different from the questions raised by the Milgram experiments. What we have to ask is whether subjects can validly consent to any magnitude of potential harm. Our analysis of that question will not be helped by looking at the Milgram experiments.

THE CONSTRUCTION OF TUSKEGEE

The Tuskegee study only ended in 1972 because the ethical concerns relating to it were printed on the front page of a prominent newspaper. The national outcry that followed was rekindled over the years and eventually resulted in a formal apology from President Clinton in 1997. Tuskegee has been heralded as a prime example of unethical and scientifically worthless research involving deliberate non-treatment, under-treatment, deception,[4] and exploitation. The study has also been labelled racist and an example of "genocide." It is now generally referred to as the "infamous Tuskegee experiment"[5], although it is not strictly an experiment. Let us look at the transformations that the study undergoes when it is used in present discussions.

Scientific Validity

Many commentators are critical of the Tuskegee study for lacking scientific validity. They take the view that the research participants suffered in vain. Today the gold standard for research is the randomised controlled trial, and the Tuskegee study did not conform to this type of design. However, compared with the standards at the time of the beginning of the trial, the design was quite common.
Other aspects of scientific validity more clearly fall short of today's standards. One commentator argues: "The research began without formal protocols, continued sloppily, and never produced substantial findings." (Rothman, 1982, p. 7). There were also numerous changes in personnel and irregular gaps between observations, weakening continuity.

Were the study to come before a research ethics committee today, it would undoubtedly be turned down on a number of grounds, including the lack of scientific validity. However, in 1930 the observational method would have been considered perfectly adequate, and the absence of a clearly defined protocol at the beginning of the study, was largely a product of the times (McDonald, 1974).
In fact, many recognise that Tuskegee was and remains an important contribution to medical knowledge (McDonald, 1974, Caplan, 1992). The study produced useful results in an area where information was desperately needed, and its scientific results are still quoted today.

It is thus difficult to sustain the claim that the study was completely without scientific validity, unless one relies on the simplifying assumption that there are only two options, a study is either valid or it is invalid. This is clearly the assumption that is used when the Tuskegee study is presented a devoid of any validity (and there is an interesting parallel here to discussions about the Nazi concentration camp experiments), but it is an extremely dubious assumption. Validity comes in

[4] For example, some participants were wrongly told that the spinal taps they received constituted treatment when really they had purely diagnostic purposes.
[5] A search on www.google.com with the search words "Tuskegee and infamous" identifies more than 2100 web-pages!

degrees, just like epistemic warrant or verisimilitude.

Lack of Informed Consent

Perhaps the most widely held criticism of the study is that it lacked procedures for informed consent. In 1930 when the study began, the importance of consent was accepted to some degree, but largely as a procedure whereby the participant assents to various medical interventions. Today, and since the Nuremberg and Helsinki Declarations, consent also involves *voluntary* assent, and adequate information. Consent was sparse in the Tuskegee study. Even where there was assent, information was inadequate and deception was used to persuade participants to undergo diagnostic spinal taps. Allegedly, some patients were not even told they had syphilis, but warned that they had 'bad blood' which was interpreted in a number of ways. Though we might accept lower standards in 1930, due to the lack of ethical guidance, the US Public Health Service cannot reasonably be excused after promulgation of the Declarations of Nuremberg and Helsinki.

It might be argued that the participants would not have consented to the spinal tap procedure without the deception, so defeating the research. Yet this is no justification; not only was the procedure itself of greater than minimal risk, but the deception undoubtedly caused harm, particularly in the form of non-treatment and undertreatment.

Non-Treatment and Undertreatment

The Tuskegee study was labelled one "of nature." It was intended to observe the course of untreated syphilis in the light of evidence that salvarsan in the 1930s was of limited efficacy and availability.

In fact the vast majority of participants received some form of treatment over time, but it was usually sporadic and was not supported by the study.[6] The doubts as to the efficacy of Salvarsan in the 1930s seem valid in the light of the 'Oslo study.' However, penicillin was of proven efficacy, and was widely available from the 1953. The Declaration of Helsinki guarantees research participants rights to the best proven treatment but debate has waged continually as to whether this is an absolute concept, or simply guarantees the patient the best *that would otherwise have been available* to him or her. Were it the case that the population was not denied treatment, but also not specially singled out for treatment, then on the latter interpretation of the Helsinki Declaration, this aspect of the ethical dilemma might appear less severe. Conversely, Rothman, 1982 argues that the study should not even be labelled 'a study of nature' because it relied on social rather than biological conditions. Thus, had there been no known treatment for syphilis, a study of the natural course of the disease might have been ethical (had informed consent been obtained). Had the treatment been of such limited efficacy that a trial was ordered to compare the natural course of the disease with medically treated syphilis, this too may have been satisfactory. However, Tuskegee involved the use of participants who were socially deprived of the treatment and thus Rothman maintains that is was unethical to use them in the study of the course of the disease.

On the other hand, Kampmeier, 1972 &1974 challenges the presumption that penicillin was the best *proven* treatment. In fact, he claims, penicillin was not of proven efficacy in the treatment of syphilis contracted up to 20 years before.[7] His

[6] It was reluctantly reported (as evidence of treatment did not bode well for the scientific validity of a 'study of nature') by those conducting the experiment, that "approximately 96% of those examined had received some therapy other than an incidental antibiotic injection and perhaps as many as 33% had curative therapy." See Reverby, 1999, p. 9.

[7] Kampmeier 1974, p. 1352. "The natural history of syphilis being what it is, the belief that the exhibition of penicillin for treatment of these men in 1953, two or more decades after infection, would have altered

claim is that the state of medical knowledge in 1953, when penicillin became widely available stillwarranted a study of the natural course of late syphilis. Of course, the issue remains that the participants had not consented to being the individuals to test this hypothesis. However, Kampmeier has "no intention of becoming entrapped in this topic since to me the interpretation of customs and mores of a generation or two ago in today's terms is a tilting at windmills. . . . Jenner, the Hunters, Pasteur and hundreds of thousands of physicians and surgeons of the past have been unethical in terms of "informed consent."

Arguably the study stood out more because of its high media visibility and racial connotations than the breaches of informed consent and right to best proven treatment.

Exploitation of a Socially Vulnerable Group: Racism

In 1981 James Jones published a book entitled "Bad Blood" in which he focused on the issue of racism. The experiment concerned black African Americans and many have concluded that the exploitation of this group of participants was racist in nature. Brandt, 1978 points out that when the experiment began there was a belief amongst the medical profession that black Americans were suffering an evolutionary degeneration. It was thought that they were particularly sexually promiscuous, and would be likely to shun treatment for sexually transmitted disease. It is possible that the study was trying to verify a myth existent at the time, which proclaimed that there are biological differences between black and white which result in different medical problems and necessary treatments. It is also suggested by some that the study may have been trying to dispel that myth (Reverby, 1999).

The people of Tuskegee (where the syphilis rate was 40%), were originally selected by the US Public Health Service for research attempting to eradicate syphilis with continued treatment of salvarsan. Unfortunately, the Depression hit the USA and money ran out. However, as those with syphilis had been identified, it was decided that a much needed observation of late syphilis would be undertaken. The original intention was therefore to single out a population with a high rate of syphilis who had received little treatment to date.

Though it may be a valid point that issues of racism lie behind the study, it is perhaps wise to be wary of using a single case such as this to demonstrate racism in the medical research community. There is evidence of large scale distrust amongst the African American community, much of which stems from the Tuskegee study. Rothman argues:

> Case studies, like the one by Jones, can be intellectually dangerous undertakings, for unless they are informed by a wider context one cannot easily distinguish among causative considerations. . . . A larger perspective on Tuskegee reduces the contribution of race and gives the styles and mindsets of the research scientist much greater prominence. (Rothman, 1982, p. 5).

The Infamous Tuskegee Study

The current received version of "The Infamous Tuskegee Study" is as a racist, scientifically invalid, unnecessary study that was ethically unacceptable already at its inception in 1932. This can perhaps best be illustrated by quoting the Presidential apology of May 16, 1997:

> The eight men who are survivors of the syphilis study at Tuskegee are a living link to a time not so very long ago that many Americans would prefer not to remember, but we dare not forget. It was a time when our nation failed to live up to its ideals, when our nation broke the trust with our people that is the very foundation of our democracy. It is not only

---

the prognosis is without basis in fact."

> in remembering that shameful past that we can make amends and repair our nation, but it is in remembering that past that we can build a better present and a better future. And without remembering it, we cannot make amends and we cannot go forward.
> So today America does remember the hundreds of men used in research without their knowledge and consent. We remember them and their family members. Men who were poor and African American, without resources and with few alternatives, they believed they had found hope when they were offered free medical care by the United States Public Health Service. They were betrayed.
> Medical people are supposed to help when we need care, but even once a cure was discovered, they were denied help, and they were lied to by their government. Our government is supposed to protect the rights of its citizens; their rights were trampled upon. Forty years, hundreds of men betrayed, along with their wives and children, along with the community in Macon County, Alabama, the City of Tuskegee, the fine university there, and the larger African American community.
> The United States government did something that was wrong – deeply, profoundly, morally wrong. It was an outrage to our commitment to integrity and equality for all our citizens. [. . .]
> The people who ran the study at Tuskegee diminished the stature of man by abandoning the most basic ethical precepts. They forgot their pledge to heal and repair. They had the power to heal the survivors and all the others and they did not. Today, all we can do is apologize. (Remarks by the President in apology for study done in Tuskegee, 1997)

But as we have seen above, all the elements of this received view are disputed. There are many problems with the Tuskegee study and very little doubt that it was deeply flawed and ethically problematic for a variety of reasons. But there is at the same time very little doubt that it was not as flawed or as unique as the received view claims. Many of its flaws are shared with other studies from the same period of our history, and many of the ethical problems were exacerbated by decisions made many years after the start of the project. That is exactly what makes it important not to proceed to a blanket condemnation, but to continue a debate about exactly what went wrong and why. Transforming the study into a paradigm of racist, unethical science may be useful for rhetorical reasons in order to create a rallying point for all good anti-racist forces. It is, however, also dangerous because we may end up forgetting that the Tuskegee study was not unique and that the American and other governments has done similar things on numerous other occasions.

MILGRAM AND THE EXPLANATION OF EVIL ACTIONS

Paradoxically the literature also contains a positive use of the Milgram experiments, but in this segment of the literature the focus is not on the methods but on the results. Already in his first article Milgram connected his the findings of his experiments with the atrocities of the Holocaust:

> Obedience, as a determinant of behavior, is of particular relevance to our time. It has been reliably established that from 1933–45 millions of innocent persons were systematically slaughtered on command. Gas chambers were built, death camps were guarded, daily quotas of corpses were produced with the same efficiency as the manufacture of appliances. These inhumane policies may have originated in the mind of a single person, but they could only be carried out on a massive scale if a very large number of persons obeyed orders (Milgram, 1963, p. 371).

and the Milgram studies have later been invoked to understand such atrocities as the My Lai massacre, the Jonestown mass suicide and many other similar incidents (Erickson, 1968). The arguments in which the findings of the obedience studies are used have the following common structure.

1. An atrocity has occurred.
2. Most of the persons involved in perpetrating the atrocity are normal people (i.e. display no obvious psychopathology and no general tendencies to hurt others)
3. Most of these people stand in an authority relationship with someone ordering them to perform certain actions.
4. *Either* a) the authority is personally present *or* b) the authority is personally absent and the acts required are some steps removed from the actual killing

etc.
5. The results of the obedience studies explains how the atrocity could have taken place given the conjunction of 2–4, and thereby dispels the apparent paradox of the conjunction of 1 & 2.

This argumentative figure has close similarities to Hannah Arendt's "banality of evil" analysis of Adolf Eichmann (Arendt, 1963), and the fact that the first Milgram results were published just after the Eichmann trial and in the same year as Arendt's book, probably led to a "synergy" between these two modes of analysing evil actions in bureaucracies.

The question is, however, whether the findings from the obedience studies can really be used in this way unless we again use not the real and complex set of experiments but the received simplified version.

In this context it is important to note that legitimate authority comes in a variety of forms, and that the legitimation of these differs. Milgram's experimenter figure possessed expert authority, partly by being the experimenter, partly by being directly associated with the prestige of Yale University; but expert authority is different from political authority in its legitimation, and may also be different in its effects (i.e. in the way it induces obedience). The expert is "an authority" whereas the politically derived authority may only be "in authority."
In obedience studies following the Milgram paradigm it has consistently been found that the effect of the experimenter's authority decreases very sharply if he or she is not physically present. This seems to distinguish this form of authority from political (or politically derived) authority which often works (and must work) at a distance.

This distinction between the kind of authority studied in the Milgram experiments and the kind of authority possessed by public servants in democracies may limit the generalisability of the Milgram findings.

Another factor potentially limiting generalisability is the fact that these studies were performed in the laboratory. The laboratory is in itself a very special social context, with a special set of rules and this may clearly influence how far people are willing to go in following the orders of the "owner" of the laboratory. Some critics of the methodology of Milgram's studies have emphasised this aspect and claimed that the laboratory, and the perceived rules of the laboratory are so different from normal life, that subjects are willing to do almost everything that an experimenter tells them to do, no matter how meaningless or repugnant it is (Patten, 1977).

CONCLUSION
As we have tried to argue above both the Milgram and the Tuskegee experiments are extremely complex and multifaceted. When they are used in bioethical and public discussions as paradigms there is, however, a strong tendency to reduce their complexity. One problematic aspect is emphasised and other aspects of the real research project bracketed. The reason for this is that the projects are often mentioned or used in papers concerned with a certain topic, for instance deceptive research, and only appear there as exemplars of problematic deception. As we have seen above the use of deception in the actual projects is quite complicated and have even been defended by some.

The irony is thus that although morality does "…not run in smooth, logical grooves but recede and advance according to the expectations and possibilities embedded in situations both more public and politicized than hypothetical dilemmas" (Damico, 1982, p. 427) the studies of real life that teaches us this lesson are very often reduced to the bare bones of hypothetical dilemmas when used as examples

in writing about bioethics.

REFERENCES

Advisory Committee on Human Radiation Experiments (1996) *The Human Radiation Experiments—Final Report of the President's Advisory Committee.* New York: Oxford University Press.

Arendt, H. (1963) *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Viking Press.

Baumrind, D. (1964) Some Thoughts on Ethics of Research: After Reading Milgram's "Behavioral Study of Obedience." *American Psychologist* **19**, 181–196.

Bowman, J.E. (1999) Tuskegee as Metaphor. *Science* **285**, 49–50.

Brandt, A. (1978) Racism and Research: The Case of the Tuskegee Syphilis Study. *Hastings Center Report* **8**, 21–29.

Caplan, A.L. (1992) Twenty Years After: The Legacy of the Tuskegee Syphilis Study. *Hastings Center Report* **22**, 29–40.

Damico, A.J. (1982) The Sociology of Justice – Kohlberg and Milgram. *Political Theory* **10**, 409–434.

Erickson, M. (1968) The Inhumanity of Ordinary People. *International Journal of Psychiatry* **6**, 278–279.

Fairchild, A.L. and Bayer, R. (1999) Uses and Abuses of Tuskegee. *Science* **284**, 919–921.

Jones, J. (1981) *Bad Blood*. New York: Free Press.

Kampmeier, R.H. (1972) The Tuskegee Study of Untreated Syphilis *Southern Medical Journal* **65**, 1247–1251.

Kampmeier, R.H. (1974) Final Report on the 'Tuskegee Syphilis Study' *Southern Medical Journal* **67**, 1349–1353.

Kaufman, H. (1967) The Price of Obedience and the Price of Knowledge. *American Psychologist* **22**, 321–322.

Korn, J.H. (1997) *Illusions of Reality: A History of Deception in Social Psychology*. Albany, NY: State University of New York Press.

Mantell, D.M. (1971) The Potential for Violence in Germany. *Journal of Social Issues* **27**, 101–112.

McDonald, C.J. (1974) The Contribution of the Tuskegee Study to Medical Knowledge. *Journal of the National Medical Association* **66**, 1–7.

Milgram, S. (1963) Behavioral Study of Obedience. *Journal of Abnormal and Social Psychology* **67**, 371–378.

Milgram, S. (1964) Issues in the Study of Obedience: A Reply to Baumrind. *American Psychologist* **19**, 848–852.

Milgram, S. (1965) Some Conditions of Obedience and Disobedience to Authority. *Human Relations* **18**, 57–76.

Milgram, S. (1975) *Obedience to Authority – An Experimental View*. London: Tavistock.

Milgram, S. (1977) Subject Reaction: The Neglected Factor in the Ethics of Experimentation. *Hastings Center Report*, **October**, 19–23.

Miller, A.G. (1986) *The Obedience Experiments: A Case Study of Controversy in Social Science*. New York: Praeger.

Patten, S.C. (1977) Milgram's Shocking Experiments. *Philosophy* **52**, 425–440.

Remarks by the President in Apology for Study Done in Tuskegee. (1997) Press Release, May 16, 1997.

Reverby, S.M. (1999) Rethinking the Tuskegee Syphilis Study. *Nursing History Review* **7**, 3–28,

Rothman, D.J. (1982) Were Tuskegee and Willowbrook, 'Studies in Nature'? *Hastings Center Report* **12**, 5-7.